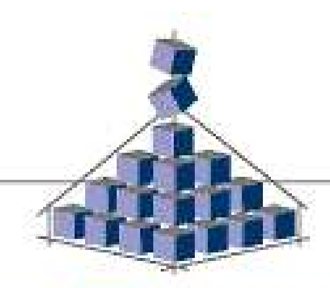




MANUAL USUARIO
PLATAFORMA IE MODULO SERVIDOR

PROYECTO:	**MODULO SERVIDOR PLATAFORMA IE**
AUTOR:	**ATTEST**
VERSION :	**1.0**
WEB :	**www.attest.es**
EMAIL :	**sw-libre@attest.es**

INDICE

1. Introducción

En este documento se presenta el manual de usuario del módulo servidor de la Plataforma de Inteligencia Estratégica.

2. CONFIGURACION DE UNA FUENTE

Para dar de alta una fuente en el sistema, debe cumplimentarse la información relacionada con: Parámetros generales, Parámetros de Búsqueda, Parámetros de recogida de resultados que se detallan a continuación.

2.1. PARÁMETROS GENERALES

Los parámetros que aquí se cumplimentan son meramente descriptivos y proporcionan información acerca de cómo es la fuente y qué información se espera obtener de ella.
Los campos que afectan directamente al usuario (ya que son éstos los que muestran a través del módulo cliente), son: "Nombre", "Categoría" e "Idioma".
El campo "Descripción" permite al administrador mantener un campo donde introducir observaciones adicionales y que no afectan al usuario.
El campo "Máximo número de resultados" permite optimizar la velocidad de recogida de datos de la fuente, limitando el número de resultados a recoger.
El campo "incidencia" avisa de la posibilidad de que dicha fuente pueda tener un problema y sea necesario depurarla. Este campo es automático, en el sentido de que si se detecta una incidencia, automáticamente se activa, para llamar la atención del Administrador. Sin embargo, éste también puede desactivarlo y activarlo a su voluntad de forma manual.
El campo "Activada" permite controlar si la fuente está funcionando en estos momentos, es decir activa. Si se desactiva, dicha fuente no estará accesible para el usuario, pero sí para el administrador, de forma que pueda editarla y reconfigurarla a su voluntad.

2.2. PARÁMETROS DE BÚSQUEDA

En esta pestaña, deben configurarse los parámetros relacionados a cómo se envía la información a la fuente. La mayor parte de esta información puede obtenerse a través de dos métodos:
- Código HTML: interpretando el código HTML de la página del formulario desde el que se realiza la búsqueda. Este codigo HTML es recomendable leerlo desde un editor de textos ASCII como el propio bloc de notas.
- Petición HTTP: observando la petición HTTP que se realiza cuando se envía la información desde el formulario. Esta información puede visualizarse a través de software específico con tal fin. Uno de estos programas es ieHTTPHeaders (http://www.blunck.info/iehttpheaders.html), que permite observar la comunicación entre el navegador cliente y el servidor (comunicación vía "cabeceras HTTP").

En primer lugar, es necesario obtener la página de búsqueda; que es la página a la que se envían los parámetros de búsqueda y la que se encarga de realizarla. Generalmente, esta página concuerda con el "action" del FORM de la página del formulario; aunque suele ser mas correcto obtener la página a la que se hace la petición HTTP a través de las cabeceras.

El campo "Método" se refiere a la forma en la que se realiza el envío de parámetros desde el formulario. Puede ser POST o GET, y se puede ver en el atributo "Method" de la etiqueta FORM de la página del formulario o en el envÍO de la cabeceras HTTP, donde se realiza un POST pagina_busqueda o GET pagina_busqueda, en función de si es un método u otro. Además, es útil saber que si se realiza una petición vía GET, los parámetros de búsqueda van a ser visibles a través de la dirección del navegador.

A continuación, es necesario leer la ayuda del formulario de la búsqueda, para conocer los operadores lógicos booleanos que éste permite (AND, OR o Frase Exacta). Una vez vistos cuáles son, se seleccionan en el campo "Operadores". Utilice el "control" para seleccionar más de uno.

El siguiente paso implica estudiar los parámetros que se envían desde el formulario. Los parámetros en HTTP tienen la forma par1=var1&par2=var2&par3=var3, etc; lo que equivale a decir que el parametro "par1" tiene valor "var1", y así sucesivamente. De todos estos parámetros se debe discernir cuáles son dinámicos y cuáles estáticos. Se llama parámetros estáticos a aquellos parámetros que son independientes del texto u operador introducido. Para ello, se introduce una cadena de búsqueda y un operador conocido y se observan los parámetros enviados. De éstos, se ven cuáles son los que están relacionados con la cadena y el operador introducido y cuáles no. Los

que no lo son, se introducen en el formato anteriormente citado en el campo de texto "Parámetros Estáticos". Los demás parámetros deben "construirse" a través de pequeños scripts de pseudoprogramación.

Por cada uno de los parámetros par1, par2, etc. que son dinámicos, se dan de alta en la fuente en la zona "Parámetros dinámicos" (introduciendo el nombre del parámetro, por ejemplo "par1" y pulsando el botón de "Añadir"). Una vez realizado, se debe asociar a cada parámetro una secuencia de instrucciones. Cuando se realice una búsqueda, el metabuscador interpretará estas instrucciones de forma secuencial y obtendrá un valor "resultado", al que asignará al parámetro asociado.

Las instrucciones que pueden interpretarse son:

- SI: condición que permite realizar una segunda instrucción si se cumple una comparación literal.
- REPLACE: permite reemplazar cadenas y subcadenas de texto por otras.
- CONCAT: permite concatenar subcadenas por delante o por detrás.
- ASSIGN: permite asignar directamente un valor literal.

Cada resultado de una instrucción se introduce en la variable "resultado", sobre la que puede operarse en la siguiente instrucción. También es necesario tener en cuenta que todos los parámetros literales de estas funciones deben ir entrecomillados.

Por ejemplo, podemos convertir texto en "texto", a través de:

CONCAT(texto,""",DELANTE)
CONCAT(resultado,""",DETRAS)

Es decir, con la primera instrucción, lo convertimos en "texto, y con la siguiente, cogemos ese resultado y le ponemos detrás otra comilla, consiguiendo "texto".

Por ejemplo, imaginemos que al hacer un envío a una fuente a la hora de buscar vino Y rosas, el navegador hace la siguiente petición: text=vino+AND+rosas&donde=todo&region=all&dias=todomes.

Se puede comprobar que el único parametro que depende del texto es "text" y los demás serian los parametros estáticos, con lo que en el cajetín de texto libre para dichos parámetros pondriamos "donde=todo&region=all&dias=todomes", y habría que añadir un nuevo parámetro dinámico "text", cuyas instrucciones permitieran pasar de "vino rosas" a "vino+AND+rosas" si fuera AND, "vino+OR+rosas" si fuera OR y ""vino rosas"" (comillas incluidas) si fuera frase exacta. Para conseguir este resultado, el conjunto de instrucciones seria:

SI OPERADOR="AND", REPLACE(texto," ","+AND+")
SI OPERADOR="OR", REPLACE(texto," ","+OR+")
SI OPERADOR="FE", CONCAT(texto,""",DELANTE)
SI OPERADOR="FE", CONCAT(texto,""",DETRAS)

Por último, debe tenerse en cuenta que automáticamente, si el método es GET, los espacios en blanco se convierten en signos "+" (lo que se denomina codificación URL), por lo que no sería necesario traducirlo en estas instrucciones.

2.3. PARÁMETROS DE RECOGIDA DE RESULTADOS

El objetivo de este apartado es definir en el sistema como se deben interpretar los resultados que devuelve el buscador para obtener los distintos resultados que se entregarán al usuario.

Al usuario se le entrega, por cada resultado, un titular, una entradilla y el enlace, por lo que habrá que definir de la página HTML que devuelve el usuario, qué es un resultado, y de éste, que es la entradilla, el titular y el enlace.

Para ello, en primer lugar el administrador debe ver el código fuente de la página de resultados que le devuelve la fuente y de éste, discernir el código HTML que obligatoriamente debe repetirse por cada resultado devuelto. Debido a que cada "porción de código HTML" correspondiente a un resultado es distinto del anterior, se utilizan expresiones regulares para delimitarlos. Las expresiones regulares permiten comparaciones del estilo "primero esta cadena, luego una cadena desconocida y lo siguiente esta cadena ".bsp;nbsp;

Por ejemplo, la expresión regular:

.+?

permite encontrar todos los enlaces con atributo name="url".

Las expresiones regulares son algo muy común en programación. Un buen tutorial se puede encontrar en: http://www.regular-expressions.info/tutorial.html. Una explicación de como funcionan éstas en Java, puede encontrase en la siguiente dirección: http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html.

Con la expresión regular correspondiente a un resultado, se debe situar entre paréntesis cada uno de los elementos definidos anteriormente: titular, entradilla y enlace. Para que el meta-buscador sepa cuál de los paréntesis se corresponde con cada uno, se debe especificar el orden en su correspondiente caja de texto. Por ejemplo "2,3,1" indica que el texto es lo que se obtiene del segundo paréntesis, la entradilla del tercero y el enlace del primero.

La opción "suprimir retornos de carro" evita incompatibilidades entre distintos sistemas y generalmente se mantendrá activado a no ser que se desee desactivarlo de forma explícita. Si se activa, el resultado HTML no tiene en cuenta los retornos de carro, lo que significa que todo el código HTML se encuentra en una misma línea, facilitando asi la obtención de la expresión regular.

Probar la fuente:

Una vez configurado todo (o la mayor parte) de la fuente, se puede comprobar si ésta funciona correctamente para ir ajustando bien los parámetros de la segunda pestaña o bien las expresiones regulares de la tercera. Esta opción simula el funcionamiento del cliente del meta-buscador, pero sólo para una fuente. Introduciendo un texto y un operador, el administrador podrá comprobar el buen funcionamiento (o no) de la fuente que acaba de dar de alta.

Una vez visto que todo funciona correctamente, el administrador debe activar la fuente (pestaña parámetros generales) para que se pueda buscar en ella.